5C

УДК 004.82

О.О. Савельев
Государственный университет информатики и искусственного интеллекта,
г. Донецк, Украина
savelyev.oleg@gmail.com

О концепции создания информационной системы интеллектуального анализа данных телекоммуникационных компаний в рамках разработки интеллектуальной системы поддержки принятия решений

В статье рассматривается проблема построения информационной системы интеллектуального анализа данных телекоммуникационных компаний. Проведено исследование и анализ современных методов, средств, технологий и идей построения систем интеллектуального анализа данных с целью определения их особенностей, а также разработана концепция собственной информационной системы интеллектуального анализа данных телекоммуникационных компаний в рамках разработки интеллектуальной системы поддержки принятия решений.

Введение

Телекоммуникационный бизнес является одной из наиболее динамически развивающихся областей современной экономики. Одним из главных преимуществ информационных корпоративных систем управления телекоммуникационными предприятиями является ведение подробного учета потребления всех видов услуг компании ее клиентами. Подобные статистические записи представляют огромный интерес с точки зрения оперативного (OLAP) и интеллектуального анализа данных (Data Mining).

Целью данной работы является исследование и анализ современных методов, средств, технологий и идей построения систем интеллектуального анализа данных, а также разработка концепции собственной информационной системы интеллектуального анализа данных телекоммуникационных компаний в рамках разработки интеллектуальной системы поддержки принятия решений.

Анализ корпоративных систем управления телекоммуникационными предприятиями

При проведении анализа основное внимание акцентировалось на интегрированных возможностях ведения учета потребления услуг компании и оперативного и интеллектуального анализа статистических данных. В подавляющем большинстве в корпоративных системах управления телекоммуникационными предприятиями для организации хранения данных применяются реляционные и объектные СУБД различных компаний, а также сопутствующее программное обеспечение. Наиболее востребованными являются продукты Oracle Application, Microsoft Business Solution Navision, SAP R/3, Lotus Notes/Domino, Foris OSS, «Парус», решения от 1С, «Атлас». Почти каждая компания старается предоставить средства интеллектуального анализа данных в рамках своих программных продуктов [1].

Фирма SAP предоставляет подсистему SAP Business Intelligence (SAP BI), предназначенную для использования в области бизнес-аналитики и поддержки принятия стратегических решений. SAP BI поддерживает концепцию хранилища данных для информации всей платформы SAP Business Suite и позволяет осуществлять ее оперативный анализ.

Крупнейший поставщик СУБД Oracle предоставляет множество продуктов в области анализа данных, среди них Oracle OLAP и Oracle Data Mining.

Фирма Microsoft в рамках своей СУБД Microsoft SQL Server 2008 предоставляет среду для создания моделей интеллектуального анализа данных Microsoft SQL Server Analysis Services, которая состоит из Business Intelligence Development Studio, SQL Server Management Studio, Microsoft SQL Server 2008 Integration Services [2], [3].

Как видно, каждая компания при разработке средств анализа данных старается создать решение, предназначенное для использования в различных предметных областях, но ориентированное на одну конкретную платформу. Поэтому создание единого комплексного средства интеллектуального анализа данных, поддерживающего множество платформ, является актуальной задачей.

Задачи интеллектуального анализа данных телекоммуникационных компаний и применение их решений

В области интеллектуального анализа данных телекоммуникационных компаний можно выделить следующие основные задачи [4]:

– анализ доходности и риска клиентов;

– выявление категорий клиентов с похожими стереотипами пользования услугами и разработка привлекательных наборов цен и услуг;

– анализ личных предпочтений отдельных клиентов, определение степени востребованности доступа к определенному контенту (сайту, серверу, абоненту телефонной связи и пр.);

– анализ зависимости использования различных видов услуг от географического местонахождения клиентов;

– анализ зависимости использования различных видов услуг от времени суток, времени года;

– анализ связей между данными (link analysis);

– защита от мошенничества.

Разработка методов решения данных задач позволит:

– осуществлять интеллектуальное управление маркетинговой политикой предприятия;

– оказывать оптимизированную персональную доставку услуг;

– обеспечивать максимально качественный уровень доступа к ресурсам и услугам организаций, для которых телекоммуникационные компании выступают в качестве посредника;

– осуществлять интеллектуальное управление распределением нагрузки на телекоммуникационное оборудование;

– решать задачи в области обеспечения предприятия запасами продукции и расходными материалами;

– обеспечивать максимальное качество предоставляемых услуг при возникновении различных отказов оборудования;

– обнаруживать случаи мошенничества и незаконных действий;

– использовать результаты интеллектуального анализа уполномоченным организациям при расследовании фактов совершения неправомерных действий.

Формирование концепции информационной системы интеллектуального анализа данных телекоммуникационных компаний

Большинство телекоммуникационных компаний пользуются различными корпоративными информационными системами, построенными на разных платформах. Поэтому при разработке систем интеллектуального анализа невозможно обеспечить большую универсальность обработки разнородной информации на уровне приложений, ориентированных на определенную платформу. Выходом из данной ситуации может служить организация платформеннонезависимой универсальной системы. Примерная схема подобной системы показана на рис. 1.

В соответствии с задачами, решаемыми СППР [5], в данной системе осуществляется ввод, хранение и анализ данных.

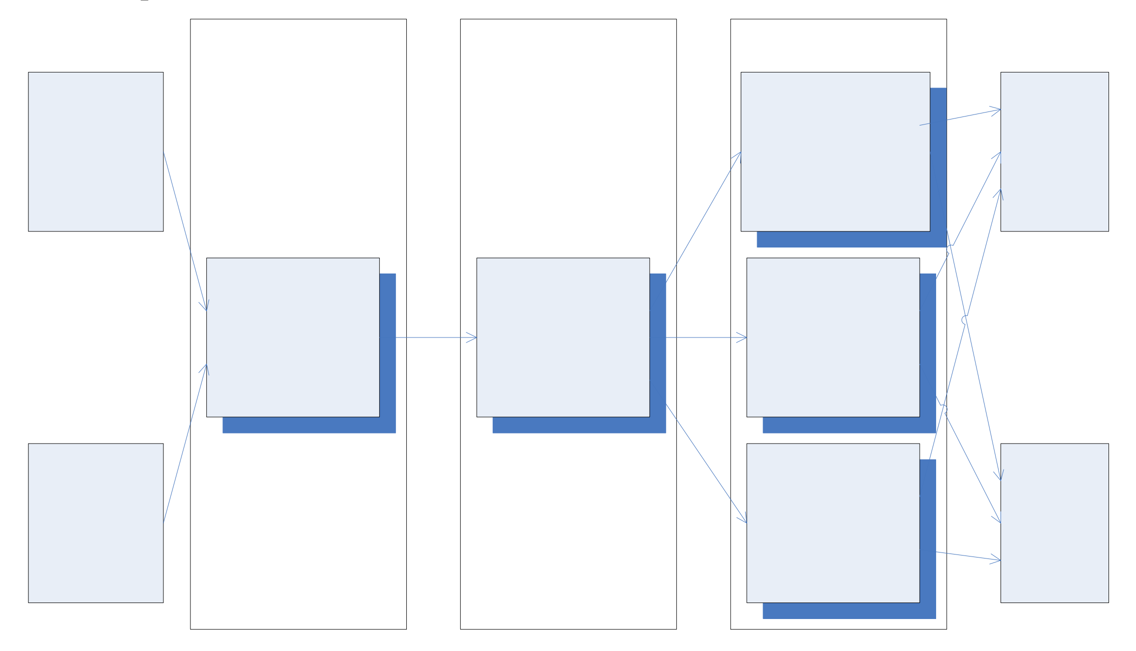


Рисунок 1 – Типовая схема универсальной СППР

Ввод данных в систему осуществляется человеком-оператором, при этом система должна предоставить удобные средства ввода, контролирующие корректность вводимых данных и выполняющие сопутствующие вычисления. Возможна реализация одновременного параллельного ввода несколькими операторами. Подсистема ввода OLTP (On-Line Transaction Processing) должна быть основана на специальных адаптивных методах и алгоритмах анализа и преобразования, позволяющих извлекать из всего потока исходных данных только те, которые отвечают используемой модели хранения данных, т.е. могут быть полезны в дальнейшем. Таким образом могут быть реализованы все функции ETL-системы (E – extraction, T – transformation, L – loading). Данный подход позволит включать в систему данные, полученные из различных источников: фрагменты разных баз данных, электронные таблицы, текст. Практика использования OLTP-систем показала неэффективность их применения.

Наиболее оптимальным средством для хранения обрабатываемой информации в подобных системах является физическое хранилище данных (Data Warehouse). Хранилище данных – предметно-ориентированный, интегрированный, неизменчивый, поддерживающий хронологию набор данных, организованный для целей поддержки принятия решений [6]. Предметная ориентация является фундаментальным отличием

Хранилища Данных от OLTP. Разные OLTP-модули могут содержать данные, описывающие одну и ту же предметную область с разных точек зрения. Решение, принятое на основе только одной точки зрения, может быть неэффективным или даже неверным. Хранилище данных позволяет интегрировать информацию, отражающую разные точки зрения на одну предметную область. Предметная ориентация позволяет также хранить в хранилище данных только те данные, которые нужны для их анализа.

В отличие от учетных систем, данные в хранилище данных после загрузки выполняют только операции чтения. Это позволяет существенно повысить скорость доступа к данным, как за счет возможной избыточности хранящейся информации, так и за счет исключения операций модификации. Цель концепции хранилищ данных – определить требования к данным, общие принципы и этапы построения хранилища данных, основные источники данных, дать рекомендации по решению потенциальных проблем, возникающих при их выгрузке, очистке, согласовании, транспортировке и загрузке.

Проблемы использования собранных данных решают подсистемы анализа. Такие подсистемы используют следующие технологии [6]: регламентированные запросы; оперативный анализ данных (OLAP); интеллектуальный анализ данных (Data Mining).

OLAP (On-Line Analytical Processing, Интерактивная аналитическая обработка) – технология оперативной аналитической обработки данных, использующая методы и средства для сбора, хранения и анализа многомерных данных в целях поддержки процессов принятия решений.

Основное назначение OLAP-систем – поддержка аналитической деятельности, произвольных запросов пользователей. Цель OLAP-анализа – проверка возникающих гипотез. В процессе принятия решений ЛПР генерирует некоторые гипотезы. Для превращения этих гипотез в законченные решения они должны быть проверены. Проверка гипотез осуществляется на основании информации об анализируемой предметной области. Как правило, наиболее удобным способом представления такой информации является зависимость между некоторыми параметрами. В процессе анализа данных OLAP позволяет объединять, просматривать и анализировать данные с точки зрения множественности измерений.

При больших объемах анализируемых данных ЛПР может не учесть всего множества возможных гипотез, в данном случае следует использовать подсистемы интеллектуального анализа данных – Data Mining.

Data Mining (Добыча данных) – исследование и обнаружение системой в сырых данных скрытых знаний, которые ранее не были известны, нетривиальны, практически полезны, доступны для интерпретации. Data Mining изучает процесс нахождения новых, действительных и потенциально полезных знаний в базах данных. Data Mining возник на пересечении систем баз данных, статистики и искусственного интеллекта. Интеллектуальный анализ включает в себя этап понимания и формулировки задачи анализа, этап подготовки данных для автоматизированного анализа, этап применения методов Data Mining и построения моделей, этап проверки построенных моделей, этап интерпретации моделей ЛПР.

Использование методов Data Mining позволяет решать задачи классификации, регрессии, поиска ассоциативных правил и кластеризации. При этом результаты решения задач могут заноситься в хранилище данных с целью дальнейшего использования и накопления опыта. Наибольший интерес в области телекоммуникаций представляет задача поиска ассоциативных правил. Решение данной задачи позволит определять закономерности происхождения событий. Одним из приложений подобных знаний является анализ данных об авариях на различных узлах сети, информация о последовательности совершения аварий может помочь в обнаружении неполадок и предупреждении новых аварий.

Немалый интерес представляет задача классификации, решение которой позволяет отделить полезную информацию от нежелательной, и задача кластеризации, решение которой позволит разбивать на кластеры географические области потребления услуг, определять социально-демографические группы потребителей, устанавливать мотивы совершения клиентами действий.

Выводы

Предложенная концепция информационной системы интеллектуального анализа данных телекоммуникационных компаний в рамках разработки интеллектуальной системы поддержки принятия решений предлагает использовать исходные данные, полученные из разных источников, объединять их в единое хранилище данных, проводить информационно-поисковый, оперативный и интеллектуальный анализ и предоставлять результаты лицам, принимающим решения.

Система, разработанная на основе данной концепции, может применяться непосредственно телекоммуникационными компаниями, концернами, сложными по структуре предприятиями для организации более эффективного управления деятельностью. Также система может найти применение при необходимости анализа узкоспециализированных слабоструктурированных, полученных из разных источников данных телекоммуникационных компаний.

Литература

1. Taniar D. Data mining and knowledge discovery technologies / David Taniar. – Hershey, NY : IGI Publiching, 2008. – 380 p.
2. Елманова Н. Введение в OLAP-технологии Microsoft / Н. Елманова, А. Федоров. – М. : Диалог-МИФИ, 2002. – 98 с.: ил.
3. Microsoft SQL Server Analysis Services. OLAP и многомерный анализ данных / [Бергер А.Б., Горбач И.В., Меломед Э.Л. и др.] ; под общ. ред. А.Б. Бергера, И.В. Горбач. – СПб. : БХВ-Петербург, 2007. – 928 с.: ил.
4. Барсегян А.А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – [2-е изд., перераб. и доп.]. – СПб. : БХВ-Петербург, 2007. – 384 с.: ил. + CD-ROM.
5. О терминах анализа. Глоссарий BI [Электронный ресурс]. – Режим доступа : http://sysdba.org.ua/faq-po-temam-/faq-po-bazam-dannyih-/o-terminah-analiza.-glossariy-bi-2.html
6. Symeonidis A.L. Agent intelligence through data mining / Andreas L. Symeonidis, Pericles A. Mitkas. – New York, NY : Springer Science+Business Media, Inc. 2005. – 215 p.

О.О. Савельєв

Про концепцію створення інформаційної системи інтелектуального аналізу даних телекомунікаційних компаній у рамках розробки інтелектуальної системи підтримки прийняття рішень

У статті розглядається проблема побудови інформаційної системи інтелектуального аналізу даних телекомунікаційних компаній. Проведено дослідження та аналіз сучасних методів, засобів, технологій та ідей побудови систем інтелектуального аналізу даних з метою визначення їх особливостей, а також розроблена концепція власної інформаційної системи інтелектуального аналізу даних телекомунікаційних компаній у рамках розробки інтелектуальної системи підтримки прийняття рішень.

O.O. Saveliev

On the Concept of an Information Data Intelligent System for Telecommunication Companies in the Development of Intellectual Decision Support Systems

The paper is devoted to the problem of construction of information data intelligent system for telecommunication companies. The investigation and analysis of modern methods, tools, technologies and ideas of construction of data mining systems to identify their characteristics were carriebont. In addition, the concept of own information data intelligent system for telecommunication companies was proposed in the development of intellectual decision support systems.

Статья поступила в редакцию 30.06.2010.